

# IDENTIFICATION AND EXPRESSION OF A NOVEL KINESIN MOTOR PROTEIN

**CROSS REFERENCE TO RELATED APPLICATIONS**  
This application claims priority to USSN 60/072,361, filed January 23, 1998, herein incorporated by reference in its entirety.

## FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with Government support under Grant No. GM35252, awarded by the National Institutes of Health, and Grant No. DMR 9612252, awarded by the National Science Foundation. The Government has certain rights in this invention.

## FIELD OF THE INVENTION

The invention provides isolated nucleic acid and amino acid sequences of TL- $\gamma$ , antibodies to TL- $\gamma$ , methods of detecting TL- $\gamma$  and screening for TL- $\gamma$  modulators using biologically active TL- $\gamma$ , and kits for screening for TL- $\gamma$  modulators.

## BACKGROUND OF THE INVENTION

The kinesin superfamily is an extended family of related microtubule motor proteins. This family is exemplified by "true" kinesin, which was first isolated from the axoplasm of squid, where it is believed to play a role in anterograde axonal transport of vesicles and organelles (*see, e.g.*, Goldstein, *Annu. Rev. Genet.* 27:319-351 (1993)). Kinesin uses ATP to generate force and directional movement associated with microtubules (from the minus to the plus end of the microtubule, hence it is a "plus-end directed" motor). Kinesin superfamily members are defined by a kinesin-like motor that is about 340 amino acids in size and shares approximately 35-45% identity (or more) with the "true" kinesin motor domain. Typically, the motor is attached to a variety of tail

domains that provide different binding activities to the various kinesin superfamily members.

The kinesin superfamily encompasses a number of families that exhibit a variety of microtubule motor functions, e.g., vesicle and organelle transport, mitotic spindle function, and meiotic spindle function. One such family is the "unc-104 family" named after unc-104 protein in *C. elegans* (Otsuka *et al.*, *Neuron* 6:13-122 (1991)). Other members of the unc-104 family include mouse Kif1A (murine homolog of unc-104) and Kif1B, and human ATSV (homolog of Kif1A) (Aizawa *et al.*, *J. Cell Biol.*, 119:1287-1296 (1992); Okada *et al.*, *Cell* 81:769-780 (1995); Nangaku *et al.*, *Cell* 79:1209-1220 (1994); and Furlong *et al.*, *Genomics* 33:421-429 (1996)). These proteins typically work as monomers, are ATP dependent, and have plus end-directed microtubule motor activity involved in fast anterograde organelle transport in neurons. Fast anterograde transport is a directional transport, typically of membranous organelles such as mitochondria, other organelles and vesicles such as synaptic vesicles, from the cell body to the tip of the axon. Members of the unc-104 family are not found in the first completely sequenced genome of *S. cerevisiae*, and on this basis it was believed that fungi did not possess members of the unc-104 motor protein family.

Accordingly, among the objects of this invention is to provide novel members of the kinesin superfamily. It is particularly an object to provide novel kinesins and nucleic acids encoding such kinesins which have anterograde axonal transport activity, including novel members of the unc-104 family, indicated by example, by sequence similarity to known kinesins having such activity. It is also an object to provide methods of use of the compounds provided herein, including methods of diagnosis, treatment of disorders related to the nervous system, treatment of disorders related to fungal infections, bioagricultural, including crop protection, and veterinary applications, and methods of identifying binding agents and modulators of the compounds provided herein.

### SUMMARY OF THE INVENTION

The present invention provides for the first time TL- $\gamma$ , an ATP-dependent, plus end-directed microtubule motor protein that is a member of the unc-104 family and the kinesin superfamily. Previously, the unc-104 family was not thought to

exist in fungi. However, the present invention surprisingly provides identification and cloning of a nucleic acid encoding TL- $\gamma$  from a hyphal fungus.

In one aspect, the invention provides an isolated nucleic acid sequence encoding a kinesin superfamily, plus end-directed microtubule motor protein, wherein the motor protein has the following properties: (i) the protein's activity includes plus end-directed microtubule motor activity; and (ii) the protein has a tail domain that has greater than 60% amino acid sequence identity to a TL- $\gamma$  tail domain as measured using a sequence comparison algorithm.

In one embodiment, the protein further specifically binds to polyclonal  
10 antibodies raised against TL- $\gamma$  of SEQ ID NO:1.

In one embodiment, the nucleic acid encodes TL- $\gamma$ . In another embodiment, the nucleic acid encodes SEQ ID NO:1. In another embodiment, the nucleic acid has a nucleotide sequence of SEQ ID NO:2.

In one embodiment, the sequence comparison algorithm is PILEUP.

15 In one embodiment, the nucleic acid is amplified by primers that selectively hybridize under stringent hybridization conditions to the same sequence as the primer set:  
5' ATGTCGGGCGGTGGAAATATC 3' (SEQ ID NO:3) and 5'  
GAATTCCTGCTTCGCTGTTTTCA 3'(SEQ ID NO:4). In another embodiment, the  
nucleic acid selectively hybridizes under stringent hybridization conditions to SEQ ID  
20 NO:2.

In one embodiment, the nucleic acid has identity to a TL- $\gamma$  derived from a hyphal fungi. In another embodiment, the nucleic acid has identity to the TL- $\gamma$  derived from *Thermomyces lanuginosus*.

In another aspect, the invention provides an expression vector comprising a  
25 nucleic acid encoding a kinesin superfamily, plus end-directed microtubule motor protein,  
wherein the motor protein has one or more of the properties described above.

In one embodiment, the invention provides a host cell transfected with the vector.

In another aspect, the invention provides an isolated kinesin superfamily, plus end-directed microtubule motor protein, wherein the protein has one or more of the properties described above.

In another aspect, the invention provides an antibody that specifically binds  
5 to TL- $\gamma$ .

In another embodiment, the invention further provides chimeric antibodies, humanized antibodies, and the nucleic acids encoding the antibodies provided herein.

15 In one embodiment, the TL- $\gamma$  specific reagent is selected from the group consisting of: TL- $\gamma$  specific antibodies, TL- $\gamma$  specific oligonucleotide primers, and TL- $\gamma$  nucleic acid probes. In another embodiment, the sample is from a plant or vertebrate, preferably a human. In another embodiment, the specific reagent is part of a gene or protein array.

In one embodiment, the method further comprises the step of isolating biologically active TL- $\gamma$  from a cell sample. In another embodiment, the biologically active TL- $\gamma$  is recombinant. In another embodiment, the biologically active TL- $\gamma$  comprises a motor, stalk or tail domain having identity to a motor, stalk or tail domain of *Thermomyces lanuginosus* TL- $\gamma$ . In another embodiment, the biologically active TL- $\gamma$  comprises an amino acid sequence of a TL- $\gamma$  motor domain of SEQ ID NO:1.

In another aspect, the invention provides a kit for screening for modulators of TL- $\gamma$ , the kit comprising; (i) a container holding biologically active TL- $\gamma$ ; and (ii) instructions for assaying for TL- $\gamma$  activity, wherein the TL- $\gamma$  activity is plus end-directed microtubule motor activity, binding activity, or ATPase activity.

5 In another aspect, the invention provides, in a computer system, a method of screening for mutations of kinesin superfamily, plus end-directed microtubule motor protein genes, the method comprising the steps of: (i) receiving input of a first nucleic acid sequence encoding a plus end-directed microtubule motor protein having a nucleotide sequence of SEQ ID NO:2 and conservatively modified versions thereof; (ii) comparing  
10 the first nucleic acid sequence with a second nucleic acid sequence having substantial identity to the first nucleic acid sequence; and (iii) identifying nucleotide differences between the first and second nucleic acid sequences.

In another aspect, the invention provides, in a computer system, a method for identifying a three-dimensional structure of kinesin superfamily, plus end-directed  
15 microtubule motor proteins, the method comprising the steps of: (i) receiving input of an amino acid sequence of at least 10 amino acids of a plus end-directed microtubule motor protein or a nucleotide sequence of at least 30 nucleotides of a gene encoding the motor protein, the protein having an amino acid sequence of SEQ ID NO:1 and conservatively modified versions thereof; and (ii) generating a three-dimensional structure of the protein  
20 encoded by the amino acid sequence.

In one aspect, the nucleic acid comprises a sequence which encodes an amino acid sequence which has one or more of the following characteristics:

greater than 60% sequence identity with SEQ ID NO:1, preferably greater than 70%, more preferably greater than 80%, more preferably greater than 90 or 95% or, in  
25 another embodiment, has 98 to 100% sequence identity with SEQ ID NO:1;

greater than 70% sequence identity with amino acids 1-357 of SEQ ID NO:1, more preferably greater than 80%, more preferably greater than 90 or 95% or, in another embodiment, has 98 to 100% sequence identity with amino acids 1-357 of SEQ ID NO:1;

greater than 60% sequence identity with amino acids 443-601 of SEQ ID NO:1,  
30 preferably greater than 70%, more preferably greater than 80%, more preferably greater than 90 or 95% or, in another embodiment, has 98 to 100% sequence identity with amino acids 443-601 of SEQ ID NO:1;

5                    In one embodiment, the nucleic acid comprises a sequence which has one or more of the following characteristics:

greater than 65% sequence identity with nucleotides 1-1071 of SEQ ID NO:2, more preferably greater than 70 or 80%, more preferably greater than 90 or 95% or, in another embodiment, has 98 to 100% sequence identity with nucleotides 1-1071 of SEQ ID NO:2;

greater than 45 or 55%, preferably greater than 65%, more preferably greater than 70 or 80%, more preferably greater than 90 or 95% or, in another embodiment, has 98 to 100% sequence identity with nucleotides 1804-2352 of SEQ ID NO:2.

In one aspect, the protein provided herein comprises an amino acid sequence which has one or more of the following characteristics:

greater than 70% sequence identity with amino acids 1-357 of SEQ ID NO:1,  
30 more preferably greater than 80%, more preferably greater than 90 or 95% or, in another  
embodiment, has 98 to 100% sequence identity with amino acids 1-357 of SEQ ID NO:1;

5 greater than 50 or 55% sequence identity with amino acids 602-784 of SEQ ID NO:1, preferably greater than 65%, more preferably greater than 70 or 80%, more preferably greater than 90 or 95% or, in another embodiment, has 98 to 100% sequence identity with amino acids 602-784 of SEQ ID NO:1.

## BRIEF DESCRIPTION OF THE DRAWINGS

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides for the first time a nucleic acid encoding TL- $\gamma$ . This protein is a member of both the kinesin superfamily of motor proteins and the unc-104 motor protein family, having ATP-dependent, plus end-directed microtubule motor activity. The surprising discovery of TL- $\gamma$  in hyphal fungi, but not non-hyphal fungi, indicates that this motor protein is likely involved in organelle transport, i.e., hyphal or axonal transport. The TL- $\gamma$  nucleic acid and protein are therefore useful in a number of diagnostic, medicinal, veterinarial, bioagricultural, and research applications. Thus, included in this invention are methods for discriminating between hyphal and non-hyphal fungi and providing means of and compositions for diagnosing hyphal fungi infections. The invention also provides methods of assaying for modulators of TL- $\gamma$  and fragments thereof as further defined herein. Such modulators include inhibitors which are useful for treating diseases caused by hyphal fungi. Such modulators are also useful for treating diseases caused by mutated TL- $\gamma$  as further defined herein, such as neurodegenerative diseases involving anterograde axonal transport. Further diagnostic, medicinal, veterinarial, bioagricultural and research applications are described below.

In one aspect, TL- $\gamma$  can be defined by having at least one or preferably more than one of the following functional and structural characteristics. Functionally,

5  
10

15  
20  
25



In one embodiment, the nucleotide sequence of TL- $\gamma$  (SEQ ID NO:2) encodes a protein of approximately 784 amino acids with a predicted molecular weight of approximately 115 kDa and a predicted range of 105 kDa to 130 kDa. TL- $\gamma$  is a member of the kinesin superfamily of motor proteins as evidenced by the sequence of its motor domain, and a member of the unc-104 motor family as evidenced by the sequence of its "stalk" domain. The predicted structure of TL- $\gamma$  consists of an amino-terminal, kinesin-like microtubule "motor" domain (approximately amino acids 1-357 of TL- $\gamma$ ); a "neck" domain (approximately amino acids 358-442 of TL- $\gamma$ ) that links the motor to the "stalk" domain, which is the unc-104 family domain (approximately amino acids 443-601 of TL- $\gamma$ ); and a "tail" domain (approximately amino acids 602 to the C-terminal end of TL- $\gamma$ ). Within the stalk domain (residues 443-601) *Thermomyces lanuginosus* TL- $\gamma$  and *C. elegans* unc-104 share 45% identity.

As defined by the sequence identity and/or hybridization characteristics described herein, the invention herein includes specific embodiments to polymorphic variants, alleles orthologs, and inter- and intra-species homologues of TL- $\gamma$ . The present invention therefore also provide polymorphic variants of TL- $\gamma$ : variant #1, in which an isoleucine residue is substituted for a valine residue at amino acid position 713; variant #2, in which a glutamic acid residue is substituted for an aspartic acid residue at amino acid position 762; and variant #3, in which a aspartic acid residue is substituted for a glutamic acid residue at amino acid position 774.

Portions of the TL- $\gamma$  nucleotide sequence may be used to identify polymorphic variants, orthologs, alleles, and homologues of TL- $\gamma$ , as well additional members of the unc-104 family of motor proteins in hyphal fungi. This identification can be made *in vitro*, e.g., under stringent hybridization conditions and sequencing, or by using the sequence information in a computer system for comparison with other nucleotide sequences. Typically, identification of polymorphic variants, orthologs, alleles, and homologues of TL- $\gamma$  is made by comparing the amino acid sequence of the tail domain (approximately amino acids 602 to the C-terminal end of TL- $\gamma$ ). Amino acid identity of approximately 50% or above, preferable 60% or above, in the tail domain typically is one characteristic that demonstrates that a protein is a TL- $\gamma$ . Similarly, the percent identity of the stalk domain is used to determine membership in the unc-104 motor protein family (approximately amino acids 443-601 of TL- $\gamma$ ). Amino acid identity

003217-9902260

of approximately 40% or above in the stalk domain typically demonstrates a protein is a member of the unc-104 family. Sequence comparison can be performed using any of the sequence comparison algorithms discussed below, with PILEUP as a preferred algorithm.

The activity of any of the peptides provided herein can be routinely confirmed by the assays provided herein such as those which assay microtubule motility, ATPase activity or binding activity. In one embodiment, polymorphic variants, alleles, and orthologs, homologues of TL- $\gamma$  are confirmed by using the microtubule motility assay described in Example II. This assay is typically used to demonstrate that a protein having about 50% or greater, preferably 60% or greater amino acid identity to the "tail" region of TL- $\gamma$  shares the same functional characteristics as TL- $\gamma$  and is therefore a species of TL- $\gamma$ . Typically, *Thermomyces lanuginosus* TL- $\gamma$  is used as a positive control in comparison to the putative TL- $\gamma$  protein to demonstrate the identification of a polymorphic variant, ortholog, allele, or homologue of TL- $\gamma$ .

TL- $\gamma$  nucleotide sequence information may also be used to construct models of the motor protein in a computer system. These models are subsequently used to identify compounds that can modulate TL- $\gamma$  or kinesin superfamily protein function.

In one embodiment, TL- $\gamma$  is found in *Thermomyces lanuginosus*, a hyphal fungus, and is predicted to exist in other hyphal fungi such as *Aspergillus* and *Neurospora*. However, TL- $\gamma$  is absent in non- or pseudo- hyphal fungi such as *Candida* and *Saccharomyces*.

The isolation of biologically active TL- $\gamma$  for the first time provides a means for assaying for modulators of this kinesin superfamily protein. Biologically active TL- $\gamma$  is useful for identifying modulators of TL- $\gamma$  or fragments thereof, unc-104 family members, and kinesin superfamily members using *in vitro* assays such as microtubule gliding assays (*see, e.g.*, Example II), ATPase assays (Kodama *et al.*, *J. Biochem.* 99:1465-1472 (1986); Stewart *et al.*, *Proc. Nat'l Acad. Sci. USA* 90:5209-5213 (1993)), and binding assays including microtubule binding assays (Vale *et al.*, *Cell* 42:39-50 (1985)). *In vivo* assays and uses are provided herein as well. Such modulators identified using biologically active TL- $\gamma$  can be used therapeutically to treat infections of hyphal fungi and neurodegenerative diseases involving defective anterograde neuronal transport. Also provided herein are methods of identifying candidate agents which bind to TL- $\gamma$  and portions thereof.

TL- $\gamma$  also provides a convenient diagnostic marker for hyphal fungi, e.g., antibodies or other probes for TL- $\gamma$  can be used *in vitro* to discriminate between hyphal fungi and non-hyphal fungi infections, e.g., between *Candida* and *Aspergillus* infections.

As further described herein, a wide variety of assays, therapeutic and diagnostic methods are provided herein which utilize the novel compounds described herein. The uses and methods provided herein, as further described below have *in vivo*, *in situ*, and *in vitro* applications, and can be used in medicinal, veterinary, bioagricultural and research based applications.

## 10 II. Definitions

As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

The terms "isolated" "purified" or "biologically pure" refer to material that is substantially or essentially free from components which normally accompany it as found in its native state. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. In particular, an isolated TL- $\gamma$  nucleic acid is separated from open reading frames that flank the TL- $\gamma$  gene and encode proteins other than TL- $\gamma$ . The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides, which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues

(Batzner *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers.

Amino acids may be referred to herein by either their commonly known  
10 three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB  
Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by  
their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration

results in the substitution of an amino acid with a chemically similar amino acid.

Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative

5 substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
  - 2) Aspartic acid (D), Glutamic acid (E);
  - 3) Asparagine (N), Glutamine (Q);
  - 4) Arginine (R), Lysine (K);
  - 10 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
  - 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).
- (see, e.g., Creighton, *Proteins* (1984)).

A "label" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include  
 15 fluorescent proteins such as green, yellow, red or blue fluorescent proteins, <sup>32</sup>P, fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available (e.g., the polypeptide of SEQ ID NO:1 can be made detectable, e.g., by incorporating a radio-label into the peptide, and used to detect antibodies  
 20 specifically reactive with the peptide).

As used herein a "nucleic acid probe or oligonucleotide" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include  
 25 natural (i.e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes  
 30 may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly

labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or subsequence.

A "labeled nucleic acid probe or oligonucleotide" is one that is bound,  
 5 either covalently, through a linker, or through ionic, van der Waals or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

"Amplification" primers are oligonucleotides comprising either natural or analogue nucleotides that can serve as the basis for the amplification of a select nucleic  
 10 acid sequence. They include, e.g., polymerase chain reaction primers and ligase chain reaction oligonucleotides.

The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, or vector, indicates that the cell, or nucleic acid, or vector, has been modified by the introduction of a heterologous nucleic acid or the alteration of a native  
 15 nucleic acid, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

A "promoter" is defined as an array of nucleic acid control sequences that  
 20 direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter  
 25 that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs  
 30 transcription of the nucleic acid corresponding to the second sequence.

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not

found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid.

An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

The term "identical" in the context of two nucleic acids or polypeptide sequences refers to the residues in the two sequences that are the same when aligned for maximum correspondence, as measured using one of the following "sequence comparison algorithms."

The phrase "substantially identical" in the context of two nucleic acids or polypeptides refers to the residues in the two sequences that have at least 50%, preferably 60%, 70%, 80%, 90% or higher identity when aligned for maximum correspondence over a domain of the protein, e.g., the stalk, motor, or tail domain of Tl- $\gamma$ , or other regions defined herein, as measured using one of the following algorithms. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection.

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences of a maximum length of 5,000. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two

to plot a dendrogram or tree representation of clustering relationships. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison, e.g., the tail, motor, or stalk domains of TL- $\gamma$ . In one example, TL- $\gamma$  was compared to other kinesin superfamily sequences (e.g., Kif1A, Kif1B and unc-104) using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. The resulting dendrogram placed TL- $\gamma$  in one cluster, with unc-104, Kif1B and Kif1A in the next most closely related cluster.

through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length  $W$  in the query sequence that either match or satisfy some positive-valued threshold score  $T$  when aligned with a word of the same length in a database sequence.  $T$  is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity  $X$  from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters  $W$ ,  $T$ , and  $X$  determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength ( $W$ ) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments ( $B$ ) of 50, expectation ( $E$ ) of 10,  $M=5$ ,  $N=-4$ , and a comparison of both strands.



The BLAST algorithm performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ( $P(N)$ ), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to an TL- $\gamma$  nucleic acid if the smallest sum probability in a comparison of the test nucleic acid to an TL- $\gamma$  nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001. Where the test nucleic acid encodes a TL- $\gamma$  polypeptide, it is considered similar to a specified TL- $\gamma$  nucleic acid if the comparison results in a smallest sum probability of less than about 0.5, and more preferably less than about 0.2.

The phrase "substantially identical" in the context of two nucleic acids or polypeptides thus typically means that a polynucleotide or polypeptide comprises a sequence that has at least about 50% identity, preferably at least about 60-80% sequence identity, and most most preferably at least about 90-95%, compared to a reference sequence using the programs described above (preferably PILEUP) using standard parameters. One indication that two nucleic acid sequences or polypeptide are substantially identical is that the polypeptide which the first nucleic acid encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is substantially identical to a second polypeptide, for example, where the two peptides differ only by a conservative substitution. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions.

The phrase "selectively hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of

5  
10

15

20

25

30

acids primarily responsible for antigen recognition. The terms variable light chain ( $V_L$ ) and variable heavy chain ( $V_H$ ) refer to these light and heavy chains respectively.

Antibodies exist e.g., as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce  $F(ab)'_2$ , a dimer of Fab which itself is a light chain joined to  $V_H-C_H1$  by a disulfide bond. The  $F(ab)'_2$  may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the  $F(ab)'_2$  dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (*see Fundamental Immunology* (Paul ed., 3d ed. 1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies (e.g., single chain Fv).

An "anti-TL- $\gamma$ " antibody is an antibody or antibody fragment that specifically binds a polypeptide encoded by the TL- $\gamma$  gene, cDNA, or a subsequence thereof.

A "chimeric antibody" is an antibody molecule in which (a) the constant region, or a portion thereof, is altered, replaced or exchanged so that the antigen binding site (variable region) is linked to a constant region of a different or altered class, effector function and/or species, or an entirely different molecule which confers new properties to the chimeric antibody, e.g., an enzyme, toxin, hormone, growth factor, drug, etc.; or (b) the variable region, or a portion thereof, is altered, replaced or exchanged with a variable region having a different or altered antigen specificity.

Humanized antibodies" refer to forms of non-human (e.g., murine) antibodies which are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin.

30 Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or

5    general, the humanized antibody will comprise substantially all of at least one, and  
typically two, variable domains, in which all or substantially all of the CDR regions  
correspond to those of a non-human immunoglobulin and all or substantially all of the FR  
regions are those of a human immunoglobulin consensus sequence. The humanized  
antibody optimally also will comprise at least a portion of an immunoglobulin constant  
10   region (Fc), typically that of a human immunoglobulin (Jones *et al.*, *Nature*, 321:522-525  
(1986); Riechmann *et al.*, *Nature*, 332:323-329 (1988); and Presta, *Curr. Op. Struct.*  
*Biol.*, 2:593-596 (1992)).

15 properties of a particular antibody to isolate, target, and/or quantify the antigen.

immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised to TL- $\gamma$  with the amino acid sequence encoded in SEQ ID NO:1 can be selected to obtain only those antibodies that are specifically immunoreactive with TL- $\gamma$  and not with other proteins, except for polymorphic variants, orthologs, alleles, and closely related homologues of TL- $\gamma$ . This selection may be achieved by subtracting out antibodies that cross react with molecules such as *C. elegans* unc-104 and mammalian Kif1. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (see, e.g., Harlow & Lane, *Antibodies, A Laboratory Manual* (1988), for a

description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

The phrase "selectively associates with" refers to the ability of a nucleic acid to "selectively hybridize" with another as defined above, or the ability of an antibody to "selectively (or specifically) bind to a protein, as defined above.

By "host cell" is meant a cell that contains an expression vector and supports the replication or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells such as plant cells, CHO, HeLa and the like. Both primary cells and tissue cultures cells are included in this definition.

"Biological sample" as used herein is a sample of biological tissue or fluid that contains TL- $\gamma$  or nucleic acid encoding TL- $\gamma$  protein. Such samples include, but are not limited to, tissue isolated from hyphal fungi or tissue suspected of being infected with hyphal fungi. Biological samples may also include sections of tissues such as frozen sections taken for histological purposes. A biological sample comprises at least one cell, preferably plant or vertebrate. Embodiments include cells obtained from a eukaryotic organism, preferably eukaryotes such as fungi, plants, insects, protozoa, birds, fish, reptiles, and preferably a mammal such as rat, mice, cow, dog, guinea pig, or rabbit, and most preferably a primate such as chimpanzees or humans.

The term “*in vivo*” refers to applications which occur in a cell while in an organism. The term “*in vitro*” refers to applications which are outside of the organism, and can be in a cell or cell free environment. “*In situ*” refers to applications which undergo a combination of environments, for example, when a cell is manipulated and then

A “therapeutic”, or the phrase “medicinal or veterinary uses” as used herein refers to a compound which is believed to be capable of modulating TL- $\gamma$  activity which can have application in both human and animal disease. Such modulators would be desirable in a number of conditions including but not limited to: neurodegenerative diseases including Alzheimer’s Disease, Parkinson’s Disease, Huntington’s Disease, Frontotemporal Dementias, Amyotrophic Lateral Sclerosis and fungal diseases such as aspergillosis, histoplasmosis, coccidiomycosis and blastomycosis.



of TL- $\gamma$ ). This domain can be used to identify other members of the unc-104 family, particularly in other hyphal fungi, through amino acid sequence identity comparison using a sequence comparison algorithm such as PILEUP. Typically, members of the unc-104 family share at least about 40% amino acid identity at the stalk domain.

The phrase "tail domain" refers to the domain of TL- $\gamma$  that can be used to identify this particular protein (approximately amino acids 602 to the C-terminal end of TL- $\gamma$ ). This domain can be used to identify TL- $\gamma$  polymorphic variants, TL- $\gamma$  alleles orthologs, and homologues of TL- $\gamma$ , through amino acid sequence identity comparison using a sequence comparison algorithm such as PILEUP.

"TL- $\gamma$ " refers to a plus end-directed microtubule motor protein found in hyphal fungi, which is a member of the kinesin superfamily of microtubule motor proteins and a member of the unc-104 family of motor proteins. TL- $\gamma$  has activity such as ATPase activity, binding activity and plus end-directed microtubule motor activity. The term TL- $\gamma$  therefore refers to polymorphic variants, alleles, orthologs, mutants, and closely related variants that have a tail domain that has greater than about 50% amino acid sequence identity, preferably about 60% amino acid sequence identity, to a TL- $\gamma$  tail domain. In addition, such polymorphic variants, orthologs, alleles, mutants, and closely related variants typically bind to polyclonal antibodies raised against SEQ ID NO:1, or antibodies raised against the tail domain of SEQ ID NO:1.

"Modulators," "inhibitors," and "activators of TL- $\gamma$ " refer to modulatory molecules identified using *in vitro* and *in vivo* assays for TL- $\gamma$  activity. Such assays include ATPase activity, microtubule gliding, microtubule depolymerizing activity, and binding activity such as microtubule binding activity. Samples or assays that are treated with a candidate agent at a test and control concentration. The control concentration can be zero. If there is a change in TL- $\gamma$  activity between the two concentrations, this change indicates the identification of a modulator. A change in activity, which can be an increase or decrease, is preferably a change of at least 20% to 50%, more preferably by at least 50% to 75%, more preferably at least 75% to 100%, and more preferably 150% to 200%, and most preferably is a change of at least 2 to 10 fold compared to a control. Additionally, a change can be indicated by a change in binding specificity or substrate..

"Biologically active" TL- $\gamma$  refers to TL- $\gamma$  that has ATPase and plus -end directed microtubule motor activity, as tested, e.g., in an ATPase assay, a microtubule

10

### III. Isolation of the gene encoding TL- $\gamma$

This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

20

25

30



## 5 B. Cloning methods for the isolation of nucleotide sequences encoding TL- $\gamma$

Amplification techniques using primers can also be used to amplify and isolate TL- $\gamma$  from DNA or RNA. The following primers can also be used to amplify a sequence of TL- $\gamma$ : 5' ATGTCGGGCGGTGGAAATATC 3' (SEQ ID NO:3) and 5' GAATTCCTGCTTCGCTGTTTTCA 3' (SEQ ID NO:4). TL- $\gamma$  can also be amplified using degenerate oligonucleotide primers (see Example I). These primers can be used, e.g., to amplify a probe of several hundred nucleotides, which is then used to screen a library for full-length TL- $\gamma$ . Alternatively, the nucleic acid for TL- $\gamma$  can be directly amplified.

TL- $\gamma$  polymorphic variants, orthologs, alleles, and homologues that are substantially identical to TL- $\gamma$  can be isolated using TL- $\gamma$  nucleic acid probes and oligonucleotides derived e.g., from SEQ ID NO:2, under stringent hybridization conditions, by screening libraries. Alternatively, expression libraries can be used to clone TL- $\gamma$  and TL- $\gamma$  polymorphic variants, orthologs, alleles, and homologues, by detecting expressed homologues immunologically with antisera or purified antibodies made against TL- $\gamma$  that also recognize and selectively bind to the TL- $\gamma$  homologue.

Appropriate primers and probes for identifying the gene encoding TL- $\gamma$  in other species such as other hyphal fungi and organisms listed herein such as animals and humans are generated from comparisons of the sequences provided herein (SEQ ID

NOS:1 and 2 and polymorphic variants #1-3). As described above, antibodies can be used to identify TL- $\gamma$  homologues. For example, antibodies made to the motor domain of TL- $\gamma$ , the tail domain of TL- $\gamma$ , the stalk domain of TL- $\gamma$ , or to the whole protein are useful for identifying TL- $\gamma$  homologues. For identification of other unc-104 family members in hyphal fungi, preferably the stalk region is used to generate primers and probes.

To make a cDNA library, one should choose a source that is rich in the mRNA of choice, e.g., TL- $\gamma$ . The mRNA is then made into cDNA using reverse transcriptase, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. Methods for making and screening cDNA libraries are well known (*see, e.g., Gubler & Hoffman, Gene 25:263-269 (1983); Sambrook et al., supra; Ausubel et al., supra*).

For a genomic library, the DNA is extracted from the tissue and either mechanically sheared or enzymatically digested to yield fragments of about 12-20 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged *in vitro*. Recombinant phage are analyzed by plaque hybridization as described in Benton & Davis, *Science* 196:180-182 (1977). Colony hybridization is carried out as generally described in Grunstein *et al., Proc. Natl. Acad. Sci. USA., 72:3961-3965 (1975)*.

An alternative method of isolating TL- $\gamma$  nucleic acid and its homologues combines the use of synthetic oligonucleotide primers and amplification of an RNA or DNA template (*see U.S. Patents 4,683,195 and 4,683,202; PCR Protocols: A Guide to Methods and Applications (Innis et al., eds, 1990)*). Methods such as polymerase chain reaction (PCR) and ligase chain reaction (LCR) can be used to amplify nucleic acid sequences of TL- $\gamma$  directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. Degenerate oligonucleotides can be designed to amplify TL- $\gamma$  homologues using the sequences provided herein. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of TL- $\gamma$  encoding mRNA in physiological samples, for nucleic acid sequencing, or for other

purposes. Genes amplified by the PCR reaction can be purified from agarose gels and cloned into an appropriate vector.

Gene expression of TL- $\gamma$  can also be analyzed by techniques known in the art, e.g., reverse transcription and amplification of mRNA, isolation of total RNA or poly A+ RNA, northern blotting, dot blotting, *in situ* hybridization, RNase protection and the like.

Synthetic oligonucleotides can be used to construct recombinant TL- $\gamma$  genes for use as probes or for expression of protein. This method is performed using a series of overlapping oligonucleotides usually 40-120 bp in length, representing both the sense and nonsense strands of the gene. These DNA fragments are then annealed, ligated and cloned. Alternatively, amplification techniques can be used with precise primers to amplify a specific subsequence of the TL- $\gamma$  gene. The specific subsequence is then ligated into an expression vector.

The gene for TL- $\gamma$  is typically cloned into intermediate vectors before transformation into prokaryotic or eukaryotic cells for replication and/or expression. These intermediate vectors are typically prokaryote vectors or shuttle vectors.

### C. Expression in prokaryotes and eukaryotes

To obtain high level expression of a cloned gene, such as those cDNAs encoding TL- $\gamma$ , it is important to construct an expression vector that contains a strong promoter to direct transcription, a transcription/translation terminator, and if for a nucleic acid encoding a protein, a ribosome binding site for translational initiation. Suitable bacterial promoters are well known in the art and described, e.g., in Sambrook *et al.* and Ausubel *et al.* Bacterial expression systems for expressing the TL- $\gamma$  protein are available in, e.g., *E. coli*, *Bacillus sp.*, and *Salmonella* (Palva *et al.*, *Gene* 22:229-235 (1983); Mosbach *et al.*, *Nature* 302:543-545 (1983). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available. The pET23 expression system (Novagen) is a preferred prokaryotic expression system.

The promoter used to direct expression of a heterologous nucleic acid depends on the particular application. The promoter is preferably positioned about the same distance from the heterologous transcription start site as it is from the transcription

start site in its natural setting. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the TL- $\gamma$  encoding nucleic acid in host cells. A typical expression cassette thus contains a promoter operably linked to the nucleic acid sequence encoding TL- $\gamma$  and signals required for efficient polyadenylation of the transcript, ribosome binding sites, and translation termination. The nucleic acid sequence encoding TL- $\gamma$  may typically be linked to a cleavable signal peptide sequence to promote secretion of the encoded protein by the transformed cell. Such signal peptides would include, among others, the signal peptides from tissue plasminogen activator, insulin, and neuron growth factor, and juvenile hormone esterase of *Heliothis virescens*. Additional elements of the cassette may include enhancers and, if genomic DNA is used as the structural gene, introns with functional splice donor and acceptor sites.

In addition to a promoter sequence, the expression cassette should also contain a transcription termination region downstream of the structural gene to provide for efficient termination. The termination region may be obtained from the same gene as the promoter sequence or may be obtained from different genes.

The particular expression vector used to transport the genetic information into the cell is not particularly critical. Any of the conventional vectors used for expression in eukaryotic or prokaryotic cells may be used. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and fusion expression systems such as GST and LacZ. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, e.g., c-myc.

Expression vectors containing regulatory elements from eukaryotic viruses are typically used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein Bar virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A<sup>+</sup>, pMTO10/A<sup>+</sup>, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 later promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.



#### IV. Purification of TL- $\gamma$

Either naturally occurring or recombinant TL- $\gamma$  can be purified for use in functional assays. Naturally occurring TL- $\gamma$  is purified, e.g., from *Thermomyces* and any other source of a TL- $\gamma$  homologue, such as other animals, plants, hyphal fungi

5 (Aspergillus, etc.). Recombinant TL- $\gamma$  is purified from any suitable expression system.

TL- $\gamma$  may be purified to substantial purity by standard techniques, including selective precipitation with such substances as ammonium sulfate; column chromatography, immunopurification methods, and others (*see, e.g.,* Scopes, *Protein Purification: Principles and Practice* (1982); U.S. Patent No. 4,673,641; Ausubel *et al.*,  
10 *supra*; and Sambrook *et al., supra*).

A number of procedures can be employed when recombinant TL- $\gamma$  is being purified. For example, proteins having established molecular adhesion properties can be reversibly fused to TL- $\gamma$ . With the appropriate ligand, TL- $\gamma$  can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The  
15 fused protein is then removed by enzymatic activity. Finally TL- $\gamma$  could be purified using immunoaffinity columns. A preferred purification column is an Ni-NTA column.

##### A. Purification of TL- $\gamma$ from recombinant bacteria

Recombinant proteins are expressed by transformed bacteria in large  
20 amounts, typically after promoter induction; but expression can be constitutive. Promoter induction with IPTG is a preferred method of expression. Bacteria are grown according to standard procedures in the art. Fresh or frozen bacteria cells are used for isolation of protein.

Proteins expressed in bacteria may form insoluble aggregates ("inclusion  
25 bodies"). Several protocols are suitable for purification of TL- $\gamma$  inclusion bodies. For example, purification of inclusion bodies typically involves the extraction, separation and/or purification of inclusion bodies by disruption of bacterial cells, e.g., by incubation in a buffer of 50 mM TRIS/HCL pH 7.5, 50 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 0.1 mM ATP, and 1 mM PMSF. The cell suspension is typically lysed using 2-3 passages  
30 through a French Press. The cell suspension can also be homogenized using a Polytron (Brinkman Instruments) or sonicated on ice. Alternate methods of lysing bacteria are

09724666-113300

apparent to those of skill in the art (*see, e.g., Sambrook et al., supra; Ausubel et al., supra*).

If necessary, the inclusion bodies are solubilized, and the lysed cell suspension is typically centrifuged to remove unwanted insoluble matter. Proteins that formed the inclusion bodies may be renatured by dilution or dialysis with a compatible buffer. Suitable solvents include, but are not limited to urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents which are capable of solubilizing aggregate-forming proteins, for example SDS (sodium dodecyl sulfate), 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins, accompanied by a lack of immunogenicity and/or activity. Although guanidine hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of immunologically and/or biologically active protein. Other suitable buffers are known to those skilled in the art. TL- $\gamma$  is separated from other bacterial proteins by standard separation techniques, e.g., with Ni-NTA agarose resin.

Alternatively, it is possible to purify TL- $\gamma$  from bacteria periplasm. After lysis of the bacteria, when TL- $\gamma$  is exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to skill in the art. To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM MgSO<sub>4</sub> and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard separation techniques well known to those of skill in the art.

00221-9994260

## *B. Standard Protein Separation Techniques For Purifying TL- $\gamma$*

### Solubility Fractionation

Often as an initial step, particularly if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol includes adding saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This concentration will precipitate the most hydrophobic of proteins. The precipitate is then discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, either through dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

### Size Differential Filtration

The molecular weight of TL- $\gamma$  can be used to isolated it from proteins of greater and lesser size using ultrafiltration through membranes of different pore size (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

### Column Chromatography

TL- $\gamma$  can also be separated from other proteins on the basis of its size, net surface charge, hydrophobicity, and affinity for ligands. In addition, antibodies raised



5

In addition to the detection of TL- $\gamma$  genes and gene expression using nucleic acid hybridization technology, one can also use immunoassays to detect TL- $\gamma$ . Immunoassays can be used to qualitatively or quantitatively analyze TL- $\gamma$ . A general overview of the applicable technology can be found in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988).

Methods of producing polyclonal and monoclonal antibodies that react specifically with TL- $\gamma$  are known to those of skill in the art (*see, e.g.*, Coligan, *Current Protocols in Immunology* (1991); Harlow & Lane, *supra*; Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed. 1986); and Kohler & Milstein, *Nature* 256:495-497 (1975). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors, as well as preparation of polyclonal and monoclonal antibodies by immunizing rabbits or mice (*see, e.g.*, Huse *et al.*, *Science* 246:1275-1281 (1989); Ward *et al.*, *Nature* 341:544-546 (1989)).

A number of TL- $\gamma$  comprising immunogens may be used to produce antibodies specifically reactive with TL- $\gamma$ . For example, recombinant TL- $\gamma$  or a  
25 antigenic fragment thereof such as the motor, stalk, or tail domain, is isolated as described herein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells as described above, and purified as generally described above. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Alternatively, a synthetic peptide derived from the sequences disclosed herein and  
30 conjugated to a carrier protein can be used as an immunogen. Naturally occurring protein may also be used either in pure or impure form. The product is then injected into an

Methods of production of polyclonal antibodies are known to those of skill in the art. An inbred strain of mice (e.g., BALB/C mice) or rabbits is immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to TL- $\gamma$ . When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (*see* Harlow & Lane, *supra*).

**SECRET**

Monoclonal antibodies and polyclonal sera are collected and titrated against the immunogen protein in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Typically, polyclonal antisera with a titer of  $10^4$  or greater are selected and tested for their cross reactivity against non-TL- $\gamma$  proteins or even other homologous proteins from other organisms (e.g., *C. elegans* unc-104 or mammalian Kif1), using a competitive binding immunoassay. Specific polyclonal antisera and monoclonal antibodies will usually bind with a  $K_D$  of at least about 0.1 mM, more usually at least about 1  $\mu$ M, preferably at least about 0.1  $\mu$ M or better, and most preferably, 0.01  $\mu$ M or better.



5 Biotechnology 14:845-51 (1996); Neuberger, *Nature Biotechnology* 14:826 (1996);  
Lonberg and Huszar, *Intern. Rev. Immunol.* 13:65-93 (1995).

10 the binding specificities is for the TL- $\gamma$  or a portion thereof, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

15 chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)).  
Because of the random assortment of immunoglobulin heavy and light chains, these  
hybridomas (quadromas) produce a potential mixture of ten different antibody molecules,  
of which only one has the correct bispecific structure. The purification of the correct  
molecule is usually accomplished by affinity chromatography steps. Similar procedures  
20 are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, *EMBO*  
*J.*, 10:3655-3659 (1991).

25 domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host  
30 organism. For further details of generating bispecific antibodies see, for example, Suresh *et al.*, *Methods in Enzymology*, 121:210 (1986).

5

[illegible]

15

20

30

binding immunoglobulin constant regions, such as protein A or protein G may also be used as the label agent. These proteins exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (*see generally Kronval et al., J. Immunol.* 111:1401-1406 (1973); Akerstrom *et al., J. Immunol.* 135:2589-2542

5 (1985)). The labeling agent can be modified with a detectable moiety, such as biotin, to which another molecule can specifically bind, such as streptavidin. A variety of detectable moieties are well known to those skilled in the art.

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to  
10 several hours, preferably from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, antigen, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

15

#### Non-Competitive Assay Formats

Immunoassays for detecting TL- $\gamma$  in samples may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of antigen is directly measured. In one preferred "sandwich" assay, for example, the anti-  
20 TL- $\gamma$  antibodies can be bound directly to a solid substrate on which they are immobilized. These immobilized antibodies then capture TL- $\gamma$  present in the test sample. TL- $\gamma$  is thus immobilized is then bound by a labeling agent, such as a second TL- $\gamma$  antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second  
25 antibody is derived. The second or third antibody is typically modified with a detectable moiety, such as biotin, to which another molecule specifically binds, e.g., streptavidin, to provide a detectable moiety.

#### Competitive assay formats

30 In competitive assays, the amount of TL- $\gamma$  present in the sample is measured indirectly by measuring the amount of a known, added (exogenous) TL- $\gamma$  displaced (competed away) from an anti-TL- $\gamma$  antibody by the unknown TL- $\gamma$  present in a

00321-9994260







bovine serum albumin (BSA), nonfat powdered milk, and gelatin are widely used with powdered milk being most preferred.

### Labels

5           The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most any label useful in such  
10   methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (e.g., DYNABEADS™), fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (e.g., <sup>3</sup>H, <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C, or <sup>32</sup>P), enzymes (e.g.,  
15   horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic beads (e.g., polystyrene, polypropylene, latex, etc.).

          The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety  
20   of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

          Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (e.g., biotin) is covalently bound to the molecule. The ligand then binds  
25   to another molecules (e.g., streptavidin) molecule, which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. The ligands and their targets can be used in any suitable combination with antibodies that recognize TL-γ, or secondary antibodies that recognize anti-TL-γ.

30           The molecules can also be conjugated directly to signal generating compounds, e.g., by conjugation with an enzyme or fluorophore. Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases,

09724666-112800



colorimetric detection, e.g., malachite green for end-point detection or coupled PK/LDH for continuous rate monitoring.

Such assays can be used to test for the activity of TL- $\gamma$  isolated from endogenous sources or recombinant sources. Furthermore, such assays can be used to test for modulators of TL- $\gamma$ . Modulators of TL- $\gamma$ , particularly inhibitors, can be used to control hyphal fungi infections, and modulators of TL- $\gamma$  can be used to treat neurodegenerative diseases involving members of the unc-104 family of motor proteins, including conditions including but not limited to: neurodegenerative diseases including Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Frontotemporal Dementias, and Amyotrophic Lateral Sclerosis. Modulators of TL- $\gamma$  are useful as pharmaceutical treatments for hyphal fungi infections of mammals, particularly humans, such as histoplasmosis (*Histoplasma capsulatum*), coccidioidomycosis (*Coccidioides immitis*), blastomycosis (*Blastomyces dermatitidis*), and aspergillosis (*Aspergillus fumigatus*, *A. flavus*, and *A. niger*). Such infections are particularly found in immunocompromised patients such as HIV-infected patients, cancer patients, and transplant patients.

Modulators of TL- $\gamma$  activity are tested using biologically active TL- $\gamma$ , e.g., thermostable *Thermomyces lanuginosus* TL- $\gamma$ . Modulation is tested using one of the *in vitro* or *in vivo* assays described herein, e.g., ATPase, microtubule gliding, and microtubule binding.

In specific embodiments, screens may be designed to first find candidate agents that can bind to TL- $\gamma$  (also referred to as TL gamma herein) proteins, and then these agents, and agents already known to bind to TL gamma may be used in assays that evaluate the ability of the candidate agent to modulate TL gamma activity. Thus, as will be appreciated by those in the art, there are a number of different assays which may be run; binding assays and activity assays.

Thus, in a preferred embodiment, the methods comprise combining a TL gamma protein and a candidate agent, and determining the binding of the candidate agent to the TL gamma protein. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e., at zero concentration or below the level of detection.

5

15

25

used, non-amino acid substituents may be used, for example to prevent or retard in vivo degradations.

In a preferred embodiment, the candidate agents are naturally occurring proteins or fragments of naturally occurring proteins. Thus, for example, cellular extracts containing proteins, or random or directed digests of proteinaceous cellular extracts, may be used. In this way libraries of procaryotic and eukaryotic proteins may be made for screening against TL gamma. Particularly preferred in this embodiment are libraries of bacterial, fungal, viral, plant and mammalian proteins, with the latter being preferred, and human proteins being especially preferred.

In a preferred embodiment, the candidate agents are peptides of from about 5 to about 30 amino acids, with from about 5 to about 20 amino acids being preferred, and from about 7 to about 15 being particularly preferred. The peptides may be digests of naturally occurring proteins as is outlined above, random peptides, or "biased" random peptides. By "randomized" or grammatical equivalents herein is meant that each nucleic acid and peptide consists of essentially random nucleotides and amino acids, respectively. Since generally these random peptides (or nucleic acids, discussed below) are chemically synthesized, they may incorporate any nucleotide or amino acid at any position. The synthetic process can be designed to generate randomized proteins or nucleic acids, to allow the formation of all or most of the possible combinations over the length of the sequence, thus forming a library of randomized candidate proteinaceous agents.

In one embodiment, the library is fully randomized, with no sequence preferences or constants at any position. In a preferred embodiment, the library is biased. That is, some positions within the sequence are either held constant, or are selected from a limited number of possibilities. For example, in a preferred embodiment, the nucleotides or amino acid residues are randomized within a defined class, for example, of hydrophobic amino acids, hydrophilic residues, sterically biased (either small or large) residues, towards the creation of cysteines, for cross-linking, prolines for SH-3 domains, serines, threonines, tyrosines or histidines for phosphorylation sites, etc., or to purines, etc.

In a preferred embodiment, the candidate agents are nucleic acids. By "nucleic acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will

- generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that may have alternate backbones, comprising, for example, phosphoramidate (Beaucage *et al.*, *Tetrahedron* 49(10):1925 (1993) and references therein; Letsinger, *J. Org. Chem.* 35:3800 (1970); Sprinzl *et al.*, *Eur. J. Biochem.* 81:579 (1977); Letsinger *et al.*, *Nucl. Acids Res.* 14:3487 (1986); Sawai *et al.*, *Chem. Lett.* 805 (1984), Letsinger *et al.*, *J. Am. Chem. Soc.* 110:4470 (1988); and Pauwels *et al.*, *Chemica Scripta* 26:141 (1986)), phosphorothioate (Mag *et al.*, *Nucleic Acids Res.* 19:1437 (1991); and U.S. Patent No. 5,644,048), phosphorodithioate (Briu *et al.*, *J. Am. Chem. Soc.* 111:2321 (1989), O-methylphosphoroamidite linkages (*see*
- 10 Eckstein, *Oligonucleotides and Analogues: A Practical Approach*), and peptide nucleic acid backbones and linkages (*see* Egholm, *J. Am. Chem. Soc.* 114:1895 (1992); Meier *et al.*, *Chem. Int. Ed. Engl.* 31:1008 (1992); Nielsen, *Nature*, 365:566 (1993); Carlsson *et al.*, *Nature* 380:207 (1996)). Other analog nucleic acids include those with positive backbones (Denpcy *et al.*, *Proc. Natl. Acad. Sci. USA* 92:6097 (1995); non-ionic
- 15 backbones (U.S. Patent Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowshi *et al.*, *Angew. Chem. Intl. Ed. English* 30:423 (1991); Letsinger *et al.*, *J. Am. Chem. Soc.* 110:4470 (1988); Letsinger *et al.*, *Nucleoside & Nucleotide* 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Eds. Sanghui and Cook; Mesmaeker *et al.*,
- 20 *Bioorganic & Medicinal Chem. Lett.* 4:395 (1994); Jeffs *et al.*, *J. Biomolecular NMR* 34:17 (1994); *Tetrahedron Lett.* 37:743 (1996)) and non-ribose backbones, including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Sanghui and Cook. Nucleic acids containing one or more carbocyclic sugars are also
- 25 included within the definition of nucleic acids (*see* Jenkins *et al.*, *Chem. Soc. Rev.* (1995) pp. 169-176). Several nucleic acid analogs are described in Rawls, *C & E News*, June 2, 1997 page 35. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of additional moieties such as labels, or to increase the stability and half-life of such molecules in physiological environments. In addition, mixtures of
- 30 naturally occurring nucleic acids and analogs can be made. Alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made. The nucleic acids may be single stranded or double stranded, as

5 hypoxanthine, isocytosine, isoguanine, etc.

10 In a preferred embodiment, the candidate agents are organic chemical  
moieties, a wide variety of which are available in the literature.

Generally, in a preferred embodiment of the methods herein, the TL gamma protein or the candidate agent is non-diffusibly bound to an insoluble support having isolated sample receiving areas (e.g. a microtiter plate, an array, etc.). The insoluble supports may be made of any composition to which the compositions can be bound, is readily separated from soluble material, and is otherwise compatible with the overall method of screening. The surface of such supports may be solid or porous and of any convenient shape. Examples of suitable insoluble supports include microtiter plates, arrays, membranes and beads. These are typically made of glass, plastic (e.g., polystyrene), polysaccharides, nylon or nitrocellulose, teflon™, etc. Microtiter plates and arrays are especially convenient because a large number of assays can be carried out simultaneously, using small amounts of reagents and samples. In some cases magnetic beads and the like are included. The particular manner of binding of the composition is not crucial so long as it is compatible with the reagents and overall methods of the invention, maintains the activity of the composition and is nondiffusible. Preferred methods of binding include the use of antibodies (which do not sterically block either the ligand binding site or activation sequence when the protein is bound to the support),

direct binding to "sticky" or ionic supports, chemical crosslinking, the synthesis of the protein or agent on the surface, etc. In some embodiments, microtubules can be used. Following binding of the protein or agent, excess unbound material is removed by washing. The sample receiving areas may then be blocked through incubation with

5 bovine serum albumin (BSA), casein or other innocuous protein or other moiety. Also included in this invention are screening assays wherein solid supports are not used.

In a preferred embodiment, the TL gamma protein is bound to the support, and a candidate agent is added to the assay. Alternatively, the candidate agent is bound to the support and the TL gamma protein is added. Novel binding agents include specific

10 antibodies, non-natural binding agents identified in screens of chemical libraries, peptide analogs, etc. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide variety of assays may be used for this purpose, including labeled in vitro protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, functional assays (phosphorylation assays, etc.) and

15 the like.

The determination of the binding of the candidate agent to the TL gamma protein may be done in a number of ways. In a preferred embodiment, the candidate agent is labeled, and binding determined directly. For example, this may be done by attaching all or a portion of the TL gamma protein to a solid support, adding a labeled

20 candidate agent (for example a fluorescent label), washing off excess reagent, and determining whether the label is present on the solid support. Various blocking and washing steps may be utilized as is known in the art.

In some embodiments, only one of the components is labeled. For example, the proteins (or proteinaceous candidate agents) may be labeled at tyrosine

25 positions using  $^{125}\text{I}$ , or with fluorophores. Alternatively, more than one component may be labeled with different labels; using  $^{125}\text{I}$  for the proteins, for example, and a fluorophor for the candidate agents.

In a preferred embodiment, the binding of the candidate agent is determined through the use of competitive binding assays as described in the preceding

30 section, and is explicitly not limited to antibodies.

In a preferred embodiment, the methods comprise differential screening to identify agents that are capable of modulating the activity of the TL gamma proteins. In

09724566-112800



this embodiment, the methods comprise combining a TL gamma protein and a competitor in a first sample. A second sample comprises a candidate agent, a TL gamma protein and a competitor. The binding of the competitor is determined for both samples, and a change, or difference in binding between the two samples indicates the presence of an agent capable of binding to the TL gamma protein and thus modulating its activity. That is, if the binding of the competitor is different in the second sample relative to the first sample, the agent is capable of binding to the TL gamma protein. Activity can be measured and detected by other methods as described herein such as ATPase and motility assays.

Alternatively, a preferred embodiment utilizes differential screening to identify drug candidates that bind to the native TL gamma protein, but cannot bind to modified TL gamma proteins. The structure of the TL gamma protein may be modeled, and used in rational drug design to synthesize agents that interact with that site. Drug candidates that affect TL gamma bioactivity are also identified by screening drugs for the ability to either enhance or reduce the activity of the protein.

Positive controls and negative controls may be used in the assays. Preferably all control and test samples are performed in at least triplicate to obtain statistically significant results. Incubation of all samples is for a time sufficient for the binding of the agent to the protein. Following incubation, all samples are washed free of non-specifically bound material and the amount of bound, generally labeled agent determined. For example, where a radiolabel is employed, the samples may be counted in a scintillation counter to determine the amount of bound compound.

A variety of other reagents may be included in the screening assays. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used. The mixture of components may be added in any order that provides for the requisite binding.

The proteins and nucleic acids provided herein can also be used for screening purposes wherein the protein-protein interactions of the TL gamma proteins can be identified. Genetic systems have been described to detect protein-protein interactions.

09724666-12800

The first work was done in yeast systems, namely the "yeast two-hybrid" system. The basic system requires a protein-protein interaction in order to turn on transcription of a reporter gene. Subsequent work was done in mammalian cells. See Fields *et al.*, *Nature* 340:245 (1989); Vasavada *et al.*, *PNAS USA* 88:10686 (1991); Fearon *et al.*, *PNAS USA* 89:7958 (1992); Dang *et al.*, *Mol. Cell. Biol.* 11:954 (1991); Chien *et al.*, *PNAS USA* 88:9578 (1991); and U.S. Patent Nos. 5,283,173, 5,667,973, 5,468,614, 5,525,490, and 5,637,463.

In general, two nucleic acids are transformed into a cell, where one is a "bait" such as the gene encoding TL gamma or a portion thereof, and the other encodes a test candidate. Only if the two expression products bind to one another will an indicator, such as a fluorescent protein, be expressed. Using the same system and the identified TL gamma proteins the reverse can be performed. Namely, the TL gamma proteins provided herein can be used to identify new baits, or agents which interact with TL gamma proteins. Additionally, the two-hybrid system can be used wherein a test candidate is added in addition to the bait and the TL gamma protein encoding nucleic acids to determine agents which interfere with the bait and the TL gamma protein.

Yet another assay for compounds that modulate TL- $\gamma$  activity involves computer assisted drug design, in which a computer system is used to generate a three-dimensional structure of TL- $\gamma$  based on the structural information encoded by the amino acid sequence. The amino acid sequence interacts directly and actively with a preestablished algorithm in a computer program to yield secondary, tertiary, and quaternary structural models of the protein. The models of the protein structure are then examined to identify regions of the structure that have the ability to bind to ligands. These regions are then used to identify ligands that bind to the protein.

The three-dimensional structural model of the protein is generated by inputting channel protein amino acid sequences or nucleic acid sequences encoding a channel protein into the computer system. The amino acid sequence of the channel protein is selected from the group consisting of SEQ ID NO:1 and conservatively modified versions thereof. The amino acid sequence represents the primary sequence of the protein, which encodes the structural information of the protein. The at least 10 nucleotides of the amino acid sequence (or a nucleotide sequence encoding 10 amino acids) is input into the computer system from computer readable substrates that include,



In this way, bioactive agents are identified. Compounds with pharmacological activity are able to enhance or interfere with the activity of the TL gamma protein. The compounds having the desired pharmacological activity may be administered in a physiologically acceptable carrier to a host, as previously described. The agents may be administered in a variety of ways, orally, parenterally e.g., subcutaneously, intraperitoneally, intravascularly, etc. Depending upon the manner of introduction, the compounds may be formulated in a variety of ways. The concentration of therapeutically active compound in the formulation may vary from about 0.1-100 wt.%. In the case of plants, sprays are a preferred embodiment.

30 In addition to the use of identifying modulators and the use in treatments and diagnostics, the nucleic acids provided herein can be used to provide cell and animal

models. Moreover, the nucleic acids can be used in some cases to provide antisense molecules and transgenics. Examples of such include at least the following embodiments.

Nucleotide sequences encoding a TL gamma protein can also be used to construct hybridization probes for mapping the gene which encodes that TL gamma  
5 protein and for the genetic analysis of individuals with genetic disorders. The nucleotide sequences provided herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

Nucleic acids which encode TL gamma protein can also be used to generate  
10 either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse, rat, sheep, cow) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell  
15 from which a transgenic animal develops. In one embodiment, cDNA encoding a TL gamma protein can be used to clone genomic DNA encoding a TL gamma protein in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express the desired DNA. Methods for generating transgenic animals, particularly animals such as mice or rats, have become  
20 conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for the TL gamma protein transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding a TL gamma protein introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of  
25 the desired nucleic acid. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the  
30 pathological condition.

Alternatively, homologues of the TL gamma protein can be used to construct a TL gamma protein "knock out" animal which has a defective or altered gene

0972466-11300

encoding a TL gamma protein as a result of homologous recombination between the endogenous gene encoding a TL gamma protein and altered genomic DNA encoding a TL gamma protein introduced into an embryonic cell of the animal. For example, cDNA encoding a TL gamma protein can be used to clone genomic DNA encoding a TL gamma protein in accordance with established techniques. A portion of the genomic DNA encoding a TL gamma protein can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector (*see e.g.*, Thomas and Capecchi, *Cell* 51:503 (1987) for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected (*see, e.g.*, Li *et al.*, *Cell* 69:915 (1992)). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras (*see, e.g.*, Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (1987), pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the TL gamma protein polypeptide.

It is understood that the models described herein can be varied. For example, "knock-in" models can be formed, or the models can be cell-based rather than animal models.

Nucleic acid encoding the TL gamma polypeptides, antagonists or agonists may also be used in gene therapy. In gene therapy applications, genes are introduced into cells in order to achieve *in vivo* synthesis of a therapeutically effective genetic product, for example for replacement of a defective gene. "Gene therapy" includes both conventional gene therapy where a lasting effect is achieved by a single treatment, and the administration of gene therapeutic agents, which involves the one time or repeated

administration of a therapeutically effective DNA or mRNA. Antisense RNAs and DNAs can be used as therapeutic agents for blocking the expression of certain genes *in vivo*. It has already been shown that short antisense oligonucleotides can be imported into cells where they act as inhibitors, despite their low intracellular concentrations caused by their restricted uptake by the cell membrane (Zamecnik *et al.*, *Proc. Natl. Acad. Sci. USA* 83:4143- 4146 (1986)). The oligonucleotides can be modified to enhance their uptake, e.g. by substituting their negatively charged phosphodiester groups by uncharged groups.

## VII. Diagnostic assays and kits

As described above, TL- $\gamma$  and its homologues are also a useful diagnostic tool *in vitro* for hyphal fungi infections and for discriminating between hyphal and non- or pseudo-hyphal fungi. Such assays use TL- $\gamma$  specific reagents that specifically hybridize to TL- $\gamma$  nucleic acid, such as TL- $\gamma$  probes and primers, and TL- $\gamma$  specific reagents that specifically bind to the TL- $\gamma$  protein, e.g., TL- $\gamma$  antibodies.

Nucleic acid assays for the presence of TL- $\gamma$  DNA and RNA in a sample are useful diagnostic assays. Numerous techniques are known to those skilled in the art, including Southern analysis, northern analysis, dot blots, RNase protection, S1 analysis, amplification techniques such as PCR and LCR, and *in situ* hybridization. In *in situ* hybridization, for example, the target nucleic acid is liberated from its cellular surroundings in such as to be available for hybridization within the cell while preserving the cellular morphology for subsequent interpretation and analysis. The following articles provide an overview of the art of *in situ* hybridization: Singer *et al.*, *Biotechniques* 4:230-250 (1986); Haase *et al.*, *Methods in Virology*, vol. VII, pp. 189-226 (1984); and *Nucleic Acid Hybridization: A Practical Approach* (Hames *et al.*, eds. 1987). In addition, TL- $\gamma$  protein can be detected with the various immunoassay techniques described above. The test sample is typically compared to both a positive control (e.g., a sample from *Thermomyces lanuginosus* TL- $\gamma$  or a sample expressing recombinant TL- $\gamma$ ) and a negative control (e.g., a negative sample from *Saccharomyces*).

The present invention also provides for kits for screening for modulators of TL- $\gamma$ . Such kits can be prepared from readily available materials and reagents. For example, such kits can comprise any one or more of the following materials: biologically active TL- $\gamma$ , reaction tubes, and instructions for testing TL- $\gamma$  activity. Preferably, the kit

contains biologically active *Thermomyces lanuginosus* TL- $\gamma$ . A wide variety of kits and components can be prepared according to the present invention, depending upon the intended user of the kit and the particular needs of the user. For example, the kit can be tailored for ATPase assays, microtubule gliding assays, or microtubule binding assays.

5

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

15

## EXAMPLES

The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

### 20 Example I: Cloning and expression of TL- $\gamma$

#### *A. Isolation of nucleic acid encoding TL- $\gamma$ and expression vector constructs*

Using PCR and degenerate primers, TL- $\gamma$  was amplified from *Thermomyces lanuginosus* nucleic acid. The following primers were used for amplification: GCGCGGATCCATC/TTTT/CGCC/T/ATAT/CGGT/C/A/GCAG/AAC (SEQ ID NO:5) (forward primer); GCGCGAATTCTCA/G/TGAA/G/T/CCCA/G/TGCA/C/GAG/AG/ATCA/G/T/CAC (SEQ ID NO:6) (reverse primer); and GCGCGAATTCTCA/G/TCTA/G/T/CCCA/G/TGCA/C/GAG/AG/ATCA/G/T/CAC (SEQ ID NO:7) (reverse primer). The PCR reaction used total genomic DNA from *Thermomyces lanuginosus* as the initial tissue source for amplifying a portion of the motor domain of TL- $\gamma$  using the degenerate primers. This nucleic acid sequence was then used as a probe to isolate a longer TL- $\gamma$  sequence. Both a library in lambda charon 4A (Nazar

09724666-1128000



*et al.*, *PNAS* 82:5608-5611 (1985) and a cDNA library (InVitrogen) were used for screening with the TL- $\gamma$  probe to clone a full-length TL- $\gamma$  sequence.

Clones of TL- $\gamma$  were amplified with the degenerate primers using the following conditions: (1) 10 cycles of 1 minute at 94°C, 1.5 minutes at 30°C, and 1 minute at 72°C followed by (2) 30 cycles of 1 minute at 94°C, 1 minute at 45°C, and 1 minute at 72°C. The clones were sequenced according to standard techniques. The following hybridization conditions were used to screen the genomic and cDNA libraries, using a probe from amplification with degenerate primers: 65°C overnight in 0.5 M sodium phosphate pH 7.2, 7% SDS, and 1 mM EDTA. The filter were washed 1 time for 15 minutes at 65°C in 40 mM sodium phosphate pH 7.2, 4% SDS, and 1mM EDTA and then three times at 65°C in 40 mM sodium phosphate pH 7.2, 1% SDS, 1mM EDTA. The nucleotide and amino acid sequences of TL- $\gamma$  are provided, respectively, in SEQ ID NO:2 and SEQ ID NO:1.

#### 15 *B. Expression and purification of TL- $\gamma$*

Recombinant TL- $\gamma$  was prepared in order to test its activity in a microtubule gliding assay. Recombinant TL- $\gamma$  was prepared using the pET23d expression vector, which is inducible in *E. coli* with IPTG.

First, TL- $\gamma$  of SEQ ID NO:2 was cloned into the pET expression plasmid. The TL- $\gamma$  nucleic acid was digested and ligated into pET23 (Novagen) to yield pET23-TL- $\gamma$  in the following manner. First, the region of the motor and stalk domains were amplified using PCR from a cDNA clone using primers that introduced a NdeI restriction site at the predicted translation start site from TL- $\gamma$ , and a HindIII restriction site at the 3' of this fragment. This fragment was inserted into the NdeI-HindIII sites of plasmid pET23b. Second, a HindIII-SacII fragment was excised from the genomic clone of TL- $\gamma$  (this clone was previously subcloned in pBS-SK) and ligated into the HindIII-SacII sites of pET23b to give pET23b-TL- $\gamma$ . The resulting plasmid encodes amino acids of 1-784 TL- $\gamma$ . This plasmid includes a hexahistidine tag for protein purification.

This plasmid was transformed into *E. coli* strain BL21 (DE3). A culture inoculated with a single colony was grown at 37°C in a modified LB medium (10 g bactotryptone, 5 g yeast extract, 5 g NaCl, 2 g MgSO<sub>4</sub>, 1 g casaminoacids per liter, and 200 mg ampicillin per liter) to OD 600 of -1. The cultures were allowed to cool to room

temperature and expression of fusion protein was induced with 0.5 mM IPTG at room temperature for 3-4 hours. After induction, the cells were typically used immediately to prepare TL- $\gamma$  protein. However, cells can also be pelleted and stored at -80°C or lower prior to protein isolation.

5 Cells were harvested by centrifugation approximately 4 hours after induction and resuspended in lysis buffer (50 mM Tris/HCl pH 7.5, 50 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM PMSF, and 0.1 mM ATP), and lysed by 2-3 passages through a French press. Insoluble debris was removed by a centrifugation at 35,000 rpm in TI1270 rotor for 40 minutes at 4°C. Soluble protein in the supernatant was bound in  
10 batch to 0.5 ml of NI-NTA-agarose resin (Qiagen) for 15 minutes at 4°C. The resin was placed in a column and washed with 10-20 column volumes of PEM80 buffer (80 mM PIPES pH 6.9, 1 mM EGTA, and 1 mM MgCl<sub>2</sub>) plus 10 mM imidazole. TL- $\gamma$  protein recovered by elution with PEM80 containing 100 mM imidazole. A typical yield was 1  
15 mg per 1 L culture or approximately 1% of the total soluble protein. The protein was very stable, retaining 100% activity up to 40°C after incubation for 15 minutes as measured using a microtubule dependent ATPase assay (*see, e.g., Hackney et al., J. Biol. Chem.* 264:15943-15948 (1989)). Freshly prepared protein was used to assay microtubule gliding activity, as described below.

## 20 Example II: Plus end-directed microtubule motor activity

### *A. Preparation of polarity marked microtubules and motility assay*

Microtubule gliding assays were used to determine the activity of the recombinant TL- $\gamma$  protein, e.g., whether it is a plus- or minus-end directed microtubule motor. The activity of TL- $\gamma$  was tested, as follows.

25 Taxol stabilized microtubule seeds brightly labelled with rhodamine were prepared by incubating a 1:1 ratio of rhodamine labelled bovine brain tubulin (Hyman, *et al., in Methods in Enzymology*, pp. 478-485 (Vallee, ed., 1992)) with unlabelled bovine brain tubulin at a final tubulin concentration of 2.5 mg/ml in PEM80 (80 mM Pipes pH 6.9, 1 mM EGTA, 1 mM MgCl<sub>2</sub>) containing 10% glycerol, 1  $\mu$ M taxol, 1 mM GTP at  
30 37°C for 15 minutes. This mixture was then diluted with 2.75 volumes of warm PEM80 containing 20  $\mu$ M taxol and 2 mM GTP, and sheared by 5 passes through a Hamilton syringe.

Dimly rhodamine-labelled extensions were grown from the brightly labelled seeds in PEM80 containing 1 mM GTP and 1.5 mg/ml tubulin cocktail consisting of a mixture of N-ethyl maleimide modified tubulin (Hyman, *et al.*, in *Methods in Enzymology*, pp. 478-485 (Vallee, ed., 1992)), unlabelled tubulin and rhodamine labelled tubulin at a ratio of 0.1/0.52/0.38 for 30 minutes at 37°C. The resulting suspension of polarity marked microtubules was diluted with PEM80 containing 10  $\mu$ M taxol and used to test motility.

25  $\mu$ l flow chambers prepared from cover slips sealed with an Apiezon grease, were preadsorbed with TL- $\gamma$  motor protein diluted to 0.1 mg/ml, and unbound protein removed by rinsing with 50  $\mu$ l of PEM80. A microtubule/ATP mix consisting of polarity marked microtubules in PEM80 containing 10  $\mu$ M taxol, 2 mM MgATP, and an oxygen scavenging system (0.1 mg/ml catalase, 0.03 mg/ml glucose oxidase, 10 mM glucose, 0.1%  $\beta$ -mercaptoethanol (Kishino, *et al.*, *Nature* 33:74-76 (1989)) was then flowed into the chamber.

Movement of microtubules was monitored at room temperature on a Zeiss Axioplan fluorescence microscope fitted with 63X Plan-Apochromat oil immersion objective, and a Princeton instruments cooled CCD. Automated time-lapse image acquisition and data analysis was performed using the MetaMorph software package (Universal Imaging, West Chester, PA).

#### B. TL- $\gamma$ activity

Recombinant TL- $\gamma$  protein was attached a glass coverslip using non-specific adhesion and gliding of polarity marked microtubules containing brightly fluorescent rhodamine labelled seeds near their minus ends (Howard, *et al.*, in *Motility Assays for Motor Proteins*, pp. 105-113 (Scholey, ed., 1993)) was recorded by time-lapse digital fluorescence microscopy. Microtubules moved with brightly fluorescent seeds leading, indicating that the immobilized TL- $\gamma$  protein was moving toward microtubule plus ends. No movement was observed in the absence of TL- $\gamma$ . This experiment demonstrates that TL- $\gamma$  has plus-ended microtubule motor activity.